CAPÍTULO 9

DISTRIBUCIONES DE PROBABILIDAD DISCRETAS Y CONTINUAS

1.- DISTRIBUCIONES DISCRETAS

En *R* es posible calcular valores relacionados con las distribuciones de probabilidad de las principales variables aleatorias discretas. Los nombres reservados a algunas de esas distribuciones son:

- Binomial: **`binom`**
- Hipergeométrica: **`hyper`**
- Poisson: **`pois`**
- Binomial negativa: **`nbinom`**
- Geométrica: **`geom`.**

Los nombres anteriores, sin embargo, no son sentencias de *R* que produzcan una salida válida. Es necesario anteponerles los prefijos "**d**" para la función de masa o función de probabilidad, "**p**" para la función de distribución acumulada, "**r**" para generar valores aleatorios y "**q**" para la función cuantil (inversa de la función de distribución). A continuación vemos algunos ejemplos.

- Calcular la probabilidad de que una variable aleatoria binomial de parámetros **`n=10, p=0.3`** tome el valor 4:

```
> dbinom(4,size=10,prob=0.3)
[1] 0.2001209
> dbinom(4,10,0.3)
[1] 0.2001209
```

- Probabilidad acumulada en el valor 5 (se incluye la probabilidad de este valor) de una variable aleatoria de Poisson de parámetro λ**=2:**

```
> ppois(5,2)
[1] 0.9834364
```

- Generar 10 valores aleatorios de una distribución de Poisson de parámetro 3,52:

```
> rpois(10,3.52)
 [1] 4 3 3 3 4 5 4 1 2 1
```

Es de hacer notar que cada vez que se ejecuta la sentencia anterior salen, evidentemente, valores diferentes.

- Calcular la probabilidad de conseguir 4 ases al extraer 4 cartas de una baraja (se supone que hay 8 ases). Aquí la variable aleatoria que representa el número de ases entre las 4 cartas elegidas es una variable aleatoria hipergeométrica de parámetros:

N=40; n=4; p(probabilidad inicial de éxito)=0.2

Esta distribución hipergeométrica se expresa en algunos casos como H(40,4,0.2). En otros, como es el caso de *R*, se pone en la forma H(8,32,4), siendo 8 el número de “bolas blancas”, 32 el número de “bolas negras” y 4 el número de extracciones, y éxito equivale a “bola blanca”. Por tanto, para calcular P(X=4) hacemos:

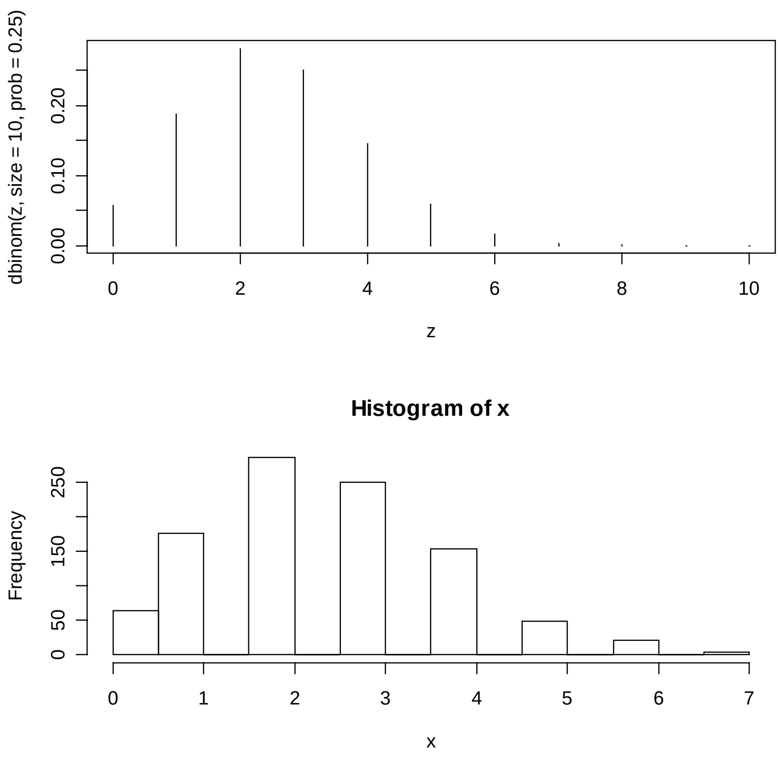
```
> dhyper(4,8,32,4)
[1] 0.0007659481
```

Este valor se podría haber obtenido de forma alternativa mediante la fórmula de la distribución hipergeométrica:

```
>choose(0.2*40,4)*choose(40-0.2*40,4-
+4)/choose(40,4)
[1] 0.0007659481
```

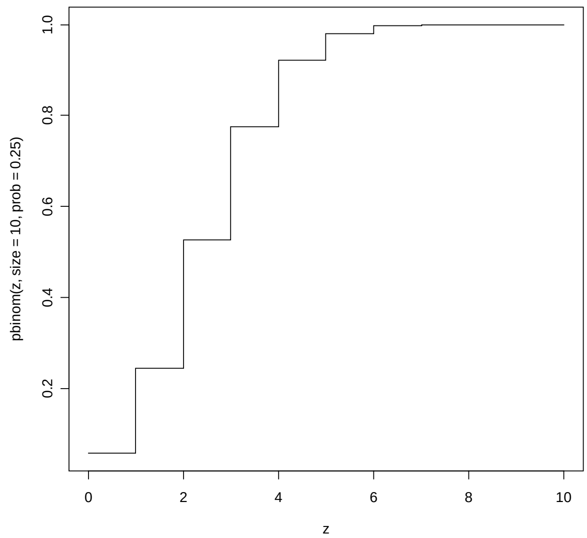
Para dibujar la función de masa de una distribución discreta debemos utilizar la función **dbinom.** A continuación aparece la correspondiente a una B(10,0.25). Debajo de esta función puede verse el histograma de 1000 valores elegidos al azar de esta misma variable aleatoria. Véase el parecido entre ambos gráficos:

```
> par(mfrow=c(2,1))
> z<-0:10
> plot(z,dbinom(z,10,0.25),type="h")
> x<-rbinom(1000,10,0.25)
> hist(x)
> par(mfrow=c(1,1))
```



Mediante la función **`pbinom`** podemos dibujar la función de distribución acumulada de una variable aleatoria. La correspondiente a la B(10,0.25) aparece a continuación. Utilizamos la opción **`type="s"`** para dibujar la función en forma de escalera.

```
> plot(z,pbinom(z,10,0.25),type="s")
```



2.- DISTRIBUCIONES CONTINUAS

En *R* también es posible calcular valores asociados a las distribuciones de probabilidad de las principales variables aleatorias continuas. Los nombres reservados a las distribuciones continuas más importantes son:

- Uniforme: **`unif`**
- Exponencial: **`exp`**
- Normal: **`norm`**
- χ^2: **`chisq`**
- t de Student: **`t`**
- F de Snedecor: **`f`**

Los nombres anteriores, igual que para las distribuciones discretas, no son sentencias de *R* que produzcan una salida válida. Es necesario anteponerles los prefijos "**`d`**" para la función de densidad, "**`p`**" para la función de distribución acumulada, "**`r`**" para generar valores aleatorios y "**`q`**" para la función cuantil (inversa de la función de distribución).

Por ejemplo, si queremos conocer la ordenada de la función de densidad de una variable aleatoria N(-2,0.2) en el punto x = -2.07, valor que por otro lado no tiene ninguna utilidad práctica, haremos:

```
> dnorm(-2.07,-2,0.2)
[1] 1.876202
```

Este resultado lo podemos obtener así mismo mediante la función de densidad de la variable aleatoria normal:

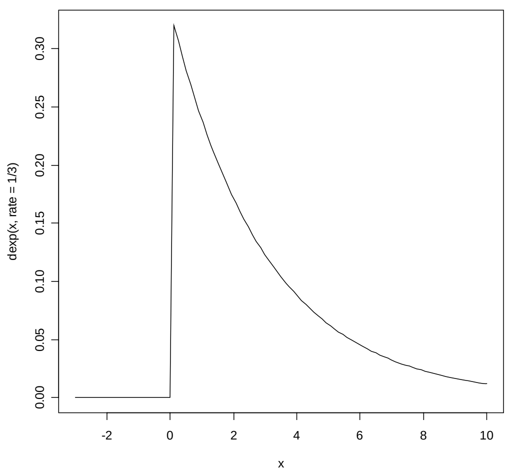
```
>exp(-(1/2)*((-2.07+2)/0.2)^2)/(0.2*sqrt(2*pi))
[1] 1.876202
```

Generemos 10 valores al azar de una distribución U(-5,3)*:*

```
> runif(10,-5,3)
 [1]    -1.8616074    -0.2718027    -3.4057091
1.3315307
 -1.6892947  2.6328703
[7] -2.8891604 -1.1120779 -0.1570343 -0.1074194
```

- En el ejemplo que se expone a continuación vamos a dibujar la función de densidad de una variable aleatoria exponencial de parámetro 3 (rate=1/3). Para ello utilizamos la función **curve.** Alternativamente podemos obtener el mismo gráfico mediante **plot**, pero definiendo previamente con el vector **z** los puntos que vamos a dibujar.

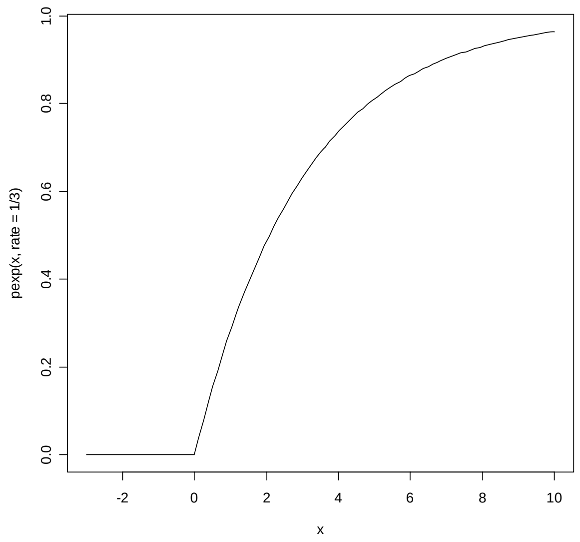
```
> curve(dexp(x,rate=1/3),from=-3,to=10)
```




```
> z<-seq(-3,10,0.1)
> plot(z,dexp(z,rate=1/3),type="l") #Con la opción type="l" conseguimos dibujar el gráfico como una línea
```

Para obtener la función de distribución acumulada volvemos a usar la función **curve**, junto con **pexp**:

```
> curve(pexp(x,rate=1/3),from=-3,to=10)
```



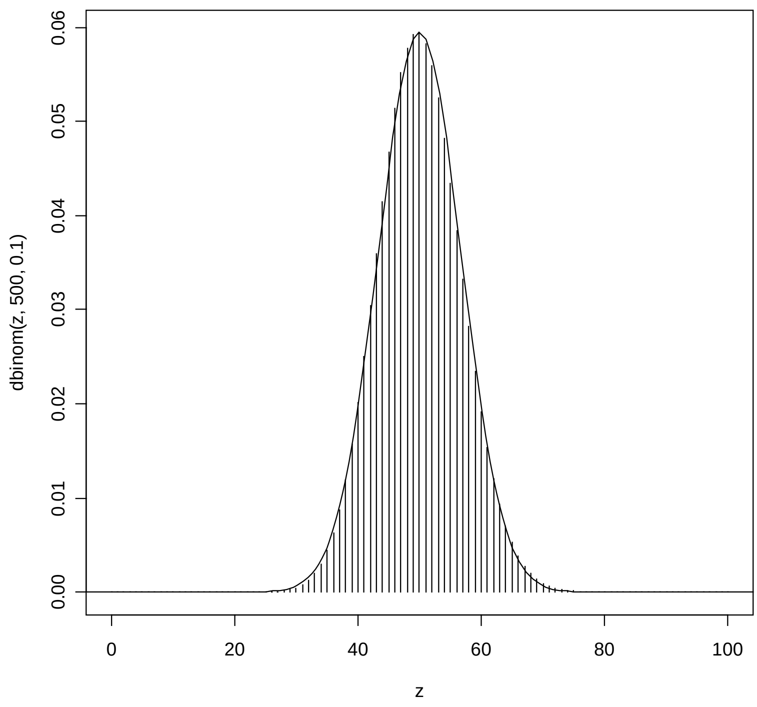
- Como se sabe por el teorema de De Moivre-Laplace, la distribución binomial puede ser aproximada, para n grande, por la distribución normal. Por ejemplo, la $B(500;0,1)$, como se cumple $500\cdot 0,1>5\ y\ 500\cdot 0,9>5$, puede ser aproximada por la $N(500\cdot 0,1;\sqrt{500\cdot 0,1\cdot 0,9})=N(50,6.71)$. Veamos en primer lugar que las probabilidades binomiales pueden ser aproximadas por la normal (se hace corrección por continuidad):

```
> pbinom(50,500,0.1)
[1] 0.5375688
> pnorm(50.5,50,6.71)
[1] 0.5297

> pbinom(60,500,0.1)-pbinom(29,500,0.1)
[1] 0.9376227
> pnorm(60.5,50,6.71)-
+ pnorm(29.5,50,6.71)
[1] 0.9400637
```

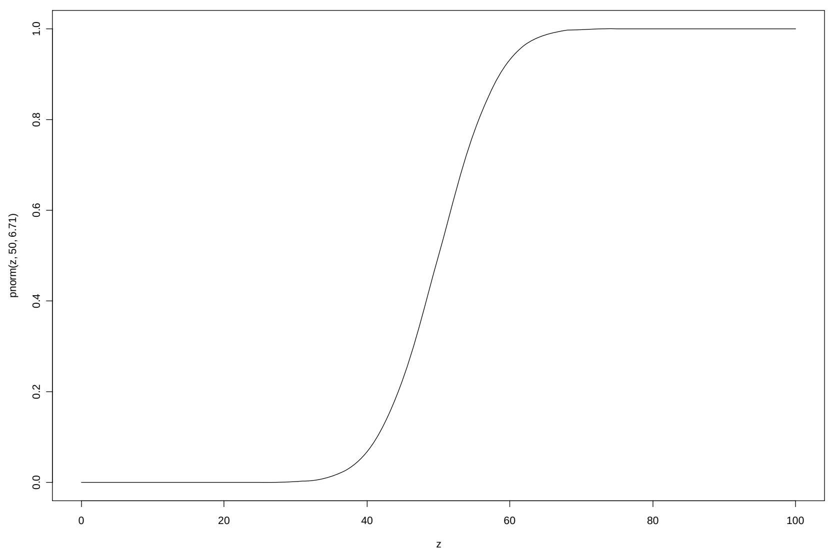
Ahora dibujamos la función de masa de la distribución B(500,0.1) y le superponemos la función de densidad de la N(50,6.71):

```
> z<-0:100
> plot(z,dbinom(z,500,0.1),type="h")
> curve(dnorm(x,50,6.71),add=T) #Utilizamos la opción add=T (TRUE) para indicarle a R que superponga este gráfico al anterior.
```



Por último vamos a dibujar la función de distribución de la variable aleatoria N(50,6.71). Como se trata de una variable aleatoria continua utilizaremos la opción **type="l"** para dibujarla como una línea, obteniendo la clásica curva en forma de *S*:

```
> plot(z,pnorm(z,50,6.71),type="l")
```



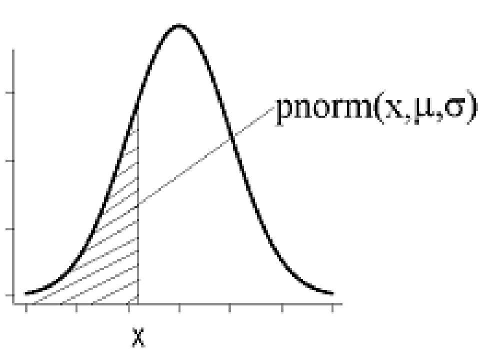
Alternativamente, podríamos construir este gráfico mediante la sentencia siguiente:

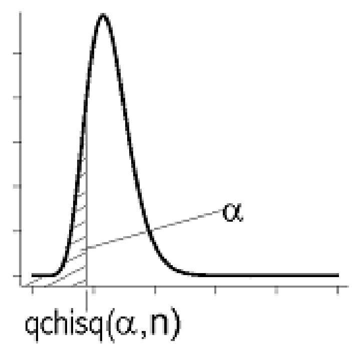
```
> curve(pnorm(x,50,6.71))
```

3.- CÓMO UTILIZAR *R* COMO ALTERNATIVA A LAS TABLAS ESTADÍSTICAS CLÁSICAS

Como se puede deducir de lo visto hasta ahora en este capítulo, el programa *R* encierra en sí mismo unas completas tablas estadísticas que, además, resultan de muy fácil manejo.

Recordemos en los gráficos siguientes el significado geométrico que tienen la función de distribución (**`pnombredeladistribución`**) y la función cuantil (**`qombredeladistribución`):**





La forma en que el programa *R* puede ser utilizado como alternativa a las tablas estadísticas clásicas se resume en el siguiente cuadro (sólo se indican las distribuciones de uso más corriente):

Distribución	*Parámetros*	*Función de masa*	*Función de distribución*	*Función cuantil*
Binomial	n, p	**dbinom(x,n,p)**	**pbinom(x,n,p)**	
Poisson	λ	**dpois(x,λ)**	**ppois(x,λ)**	
Normal	μ,σ		**pnorm(x,μ,σ)**	**qnorm(α,μ,σ)**
χ^2	n		**pchisq(x,n)**	**qchisq(α,n)**
t de Student	n		**pt(x,n)**	**qt(α,n)**
F de Snedecor	n, m		**pf(x,n,m)**	**qf(α,n,m)**

Veamos algunos ejemplos de aplicación:

```
# Calcular P(X=2) si X --> Binomial(5,0.2)
> dbinom(2,5,0.2)
[1] 0.2048

# Calcular P(X≤3) si X --> Poisson(2)
> ppois(3,2)
[1] 0.8571235

# Calcular x sabiendo que P(X≤x)=0.238 y X --> N(-
4,1)
> qnorm(0.238,-4,1)
[1] –4.712751
```

 # Calcular P(X≤34.2) si X --> χ^2(20)
> pchisq(34.2,20)
[1] 0.9751968

 # Calcular P(X≤2.14) si X --> t_{14}
> pt(2.14,14)
[1] 0.9747763

 # Calcular x si P(X≤x)=0.9 y X --> $F_{4;8}$
> qf(0.9,4,8)
[1] 2.806426